(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 0 778 351 B1

(12) EUROPEAN PATENT SPECIFICATION

(45) Date of publication and mention of the grant of the patent:
28.08.2002 Bulletin 2002/35

(51) Int Cl.^{7}: **C12Q 1/68**, B01L 7/00

(21) Application number: **96118921.4**

(22) Date of filing: **26.11.1996**

(54) Method of fingerprinting polynucleotides

Verfahren zum Fingerprinting von Polynukleotiden

Méthode de révéler des empreintes de polynucléotides

(84) Designated Contracting States:
FR GB

(30) Priority: **30.11.1995 JP 31194995**

(43) Date of publication of application:
11.06.1997 Bulletin 1997/24

(73) Proprietor: **Hitachi, Ltd.**
Chiyoda-ku, Tokyo (JP)

(72) Inventors:
- **Kambara, Hideki**
Hachioji-shi (JP)
- **Okano, Kazunori**
Shiki-shi (JP)
- **Uematsu, Chihiro**
Kokubunji-shi (JP)

(74) Representative: **Strehl Schübel-Hopf & Partner**
Maximilianstrasse 54
80538 München (DE)

(56) References cited:
EP-A- 0 630 972
EP-A- 0 649 852
EP-A- 0 701 001
WO-A-90/01064

EP 0 778 351 B1

Description

BACKGROUND OF THE INVENTION

(1) Field of the Invention

[0001] The present invention relates to a method for DNA analysis or assay.

(2) Description of the Related Art

[0002] It has been becoming popular to use DNA for diagnosis of disease. In such diagnosis of disease, (1) a DNA probe having a complementary sequence to a target DNA is prepared and a probe assay is carried out to see if this DNA probe hybridizes with the target DNA; (2) a certain region of sequence coding for a target DNA is chosen and subjected to polymerase chain reaction amplification using two DNA probes (primers), the resulting DNA fragment is read or its length is examined, and based on the DNA fragment information obtained, assay is performed; etc. The thus obtained results are utilized for diagnosis or the like. These methods are applicable to analysis or assay for a single kind of DNA, at most, a few kinds,of DNAs but not appropriate for DNA assay of a vast number of DNA fragments or total evaluation of long DNA.

[0003] However, DNAs or genomes in vivo function while interacting with each other. It is, therefore, strongly desired to collectively assess chromosomes or all DNAs contained therein. For example, in a spotlighted cDNAs among the genome-project, it has been attempted to detect the kind and amount of cDNA complementary to mRNA thereby to collectively grasp an organism, paying attention to the mechanism of DNA functioning in the organism that, where DNA functions in an organism, DNA information is first transcribed onto mRNA and a protein is synthesized based on the information to function the organism. In this attempt, cDNAs are fished out of a vital sample and the respective cDNAs are sequenced to analyze the frequency of each cDNA appearing in one tissue (body mapping).

[0004] The body mapping involves the following procedures. First, cDNA is prepared from mRNA (in a mixture of diverse cDNAs) and then cloned. E. coli containing cDNAs is spread and cultured on an agarose plate to obtain colonies, each of which contains one of the desired cDNAs. The desired cDNA is taken out and sequenced to identify the kind of cDNA. In a similar manner, a desired cDNA is taken out of each colony and sequenced, whereby the same cDNA often appears. When attention is given to one particular cDNA present in one tissue, the larger the amount of this particular cDNA, the more likely the particular cDNA corresponds to the gene strongly expressed in the tissue and as the result, the higher the frequency of the gene appearance in the colony. Accordingly, there is proposed a method for determining the frequency of cDNA appearance which comprises performing cDNA sequencing in many colonies to see how many times a particular cDNA appears in the colonies (Katsuji, Murakawa et al., Genomics, 23, 379-389 (1994)).

[0005] On the other hand, another attempt for DNA diagnosis has also been proposed, paying attention to a genome (DNAs in all chromosomes) or the entire profile of a particular chromosome. A fingerprinting technique called gene scanning, which is also called Landmark genome scanning (LGS method), involves the steps of selectively digesting DNA with 8 base cutter restriction enzyme (which digests once per 48 to 64 kbs) such as Not I, etc., binding the digested fragments to a radioisotope tag or a nucleotide labelled with fluorophore, separating the fragments by electrophoresis, then cutting the DNA fragments in a gel with a 4 base cutter restriction enzyme (which digests once per 44 to 256 bases), and subjecting the DNA fragment on the upper end of a polyacrylamide slab gel to two-dimensional electrophoresis. The thus obtained pattern is utilized as a fingerprint, thereby to comprehend the entire profile of DNA. An attempt includes use of the pattern for diagnosis, noting that a DNA pattern in normal cells is different from that of abnormal cells suffered from cancer, etc.

[0006] However, a good technique is not found so far, since it should be examined in the foregoing methods in which particular site of long DNA there is abnormality.

[0007] It is important for early detection of diseases or understanding of DNA function in cells to examine a long and large DNA or clarify the entire profile of a sample containing diverse cDNAs. As stated above, however, any good technique sufficient for the purpose has not been developed yet. According to the techniques explained hereinabove, it is necessary to determine the base sequences of very many clones. Much labor and time required make it impossible to practically apply these techniques to various samples. Conventional DNA probing is only enough to examine, at best, several to several ten kinds of DNAs in one cycle of operation but not suitable for assaying cDNAs or DNA fragments of several hundreds to thousands in one cycle of operation. In addition, the cDNA analysis methods described above are not applicable to detection of long DNA where abnormality is located.

[0008] On the other hand, the gene scanning technique can meet the foregoing requirements but encounter problems that a huge amount of enzymes are consumed in the second digestion with a restriction enzyme and in the two dimensional electropherogram, the abscissa which is a scale for length of the DNA fragment occurred in the first digestion

and the ordinate which gives some scale for length of the ultimate fragment are not always quantitative so that it is difficult to construct a database with these data.

[0009] EP-A-0 630 972 discloses a DNA analyzing method, wherein the DNA is digested with a restriction enzyme, the single stranded DNA is contacted with a DNA probe having arbitrary sequences of 2 to 6 nucleotides at the 3'-end thereof, and DNA extension only takes pleace in case the DNA matches these 2 to 6 nucleotides perfectly. The DNA fragment that binds can be sequenced subsequently.

[0010] EP-A-0 701 001 is prior art pursuant to Art. 54(3) EPC and discloses a method for separating, fractionating and analyzing DNA. This document discloses a set of primers which contain an oligomer sequence, a part of a restriction enzyme recognition sequence and a sequence of two arbitrary nucleotides.

SUMMARY OF THE INVENTION

[0011] . An object of the present invention is to overcome the foregoing problems and provide a novel fingerprinting technique, namely, a DNA assay method, which is applicable to various samples and is suitable for assaying a large number of cDHAs or DNA fragments and assaying a long DNA.

The present invention provides a method of analysis or assay for nucleotides according to claim 1.

[0012] The method of analysis or assay for nucleotides according to the present invention comprises:

(1) digesting DNA with a restriction enzyme to obtain DNA fragments;
(2) discriminating differences in sequences of 1 to 3 bases adjacent to restriction enzyme recognition sites of the DNA fragments with labeled DNA probes having all possible combinations of 1 to 3 bases out of A, C, G and T at the 3'-termini thereof;
(3) extending the labeled DNA probes by a complementary strand synthesis to .fractionate the DNA fragments into groups; and
(4) measuring the lengths of the DNA fragments which belong to said groups, or the lengths of the labeled DNA probes extended by said complementary strand synthesis;

wherein measured lengths of the DNA fragments fractionated into the groups by the differences in sequences of 1 to 3 bases adjacent to restriction enzyme recognition sites of the DNA fragments are employed as fingerprints.

[0013] In the method described above, the label may be biotin, a chemiluminescence reagent or a fluorophore; the DNA probes or the DNA fragments may be detected by detecting a fluorophore; and the labelled DNA probes may contain a set of sixteen (16) labelled DNA probes wherein at least two terminal bases are composed of substantially all of the bases species or analogues thereof.

[0014] Another characteristic feature of the present invention is that said restriction enzyme is a restriction enzyme to form a 3'-protruding end or blunt end type fragment.

Furthermore, the restriction enzyme used therein may be a restriction enzyme that can give a 5'-protruding end and the method can further include, after the step 1) described above, a step of filling the digested portions with the restriction enzyme using a DNA polymerase to form a double strand, or a step of introducing an oligomer containing the sequence of the digested portions by ligation.

[0015] In a further embodiment of the present invention the complementary strands to the DNA fragments fractionated based on a difference in two bases islabelled at least, with a fluorophore or a chemiluminescence reagent and produced by complementary strand synthesis, using labelled DNA probes having all possible combinations of arbitrary two bases at the 3' termini thereof.

[0016] The labelled DNA probes used in the method may be a set of at least sixteen (16) DNA probes in all combinations of $N_1, \ldots, N_n$ (wherein $5 \leq n \leq 2$ (7), $X_1, \ldots, X_m$ ($1 < m \leq 6$); Y_1Y_2 is any one of A, C, G and T; said labelled DNA probes, which are primers used for the complementary strand synthesis, are represented by:

5'- $N_1 \ldots N_n$, $X_1 \ldots X_m$ Y_1Y_2 - 3'

wherein $N_1 \ldots N_n$ has a substantially complementary sequence to the oligomer optionally connected to the DNA fragment; $X_1 \ldots X_m$ is substantially complementary to a part of the sequence of the portion digested with the restriction enzyme; and Y_1Y_2 is composed of any combination of two out of A, C, G and T.

[0017] Alternatively, the labelled DNA probes employed are a set of at least sixteen (16) DNA probes in all combinations of $N_1, \ldots, N_n$ (wherein $5 \leq n \leq 2$ (7), $X_1, \ldots, X_m$ ($1 \leq m \leq 6$); Y_1Y_2 is any one of A, C, G and T; Z_i ($1 \leq i \leq 3$) is a nucleotide analogue which can hybridize with a plural kinds of nucleotides; said labelled DNA probe which is a primer used for the complementary strand synthesis, is represented by:

5'- $N_1 \ldots N_n X_1 \ldots X_m Z_l \ldots Z_i Y_1 Y_2$ - 3'

wherein $N_1 \ldots N_n$ has a substantially complementary sequence to the oligomer connected to the DNA fragment; $X_1 \ldots X_m$ is substantially complementary to a part of the sequence of the portion digested with the restriction enzyme; and Y_1Y_2 is composed of any combination of two out of A, C, G and T.

[0018] Any one of the nucleotides of N_1, ..., N_n is labelled, at least, with biotin or a fluorophore.

[0019] The DNA digested with a restriction enzyme may be the DNA obtained by 1) a step of collecting mRNA from a tissue sample to prepare a mRNA-cDNA complementary pair, and 2) a step of preparing a double stranded DNA from the mRNA-cDNA complementary pair, thereby to give information on the kind and amount of mRNA.

[0020] The method of the present invention can be employed in a DNA analyzer or assay for nucleotides, wherein a plurality of the reaction vessels and a chemical reaction vessel for performing a chemical reaction While controlling temperature are provided and a set of at least sixteen (16) labelled DNA probes are separately charged in a plurality of the reaction vessels, respectively, and each of the DNA fragments charged respectively in a plurality of the reaction vessels is simultaneously reacted with the labelled DNA probe.

[0021] Even though many DNA fragments to be analyzed are present in a sample, each DNA fragment has a different sequence from that adjacent to a recognition sequence in most cases. Therefore, DNA fragments are initially fractionated into groups depending upon the difference in sequence adjacent to the recognition sequence. This grouping operation is repeated until the number of DNA fragments in each group is reduced to such a level that does not cause any inconvenience for fractionation of the DNA fragments by gel electrophoresis. Finally, the DNA fragments in each group are determined by gel electrophoresis with a length of each DNA fragment.

[0022] The DNA fragments in a sample which are obtained by digesting DNA with a restriction enzyme are fractionated and classified by complementary strand synthesis, using a set of sixteen (16) DNA probes having the two terminal bases in all combinations. The complementary strand synthesis proceeds when the two terminal bases fully hybridize with the target DNA but if not, does not proceed. The DNA strand extended by the complementary strand synthesis has substantially the same length as that of the intact DNA strand (DNA fragment obtained by digestion with a restriction enzyme). When it is so designed that a label is taken up into the DNA strand extended by the complementary strand synthesis to make thelabelled DNA strand is detectable distinguishably from other DNA strands extended by complementary strand synthesis, the terminal two bases are discriminated to determine the length of the intact DNA strand.

[0023] Where a single fractionation is insufficient for discriminating the DNA fragments obtained by digesting DNA in a sample with a restriction enzyme, biotin is tagged to DNA probe and the extended strand alone is fished out. Again the same fractionation is applied to the fished DNA fragment (extended DNA strand by complementary strand synthesis). In this case, a DNA probe employed is the one attached to the complementary strand at the 3' terminus (which corresponds to the 5' terminus of intact DNA) so that different classification from the first operation may be carried out.

[0024] That is, according to the present invention, information on lengths of many DNA fragments obtained by digesting DNA in a sample with a restriction enzyme is not utilized as fingerprints. Instead, the DNA fragments obtained by digestion with a restriction enzyme are fractionated by the terminal base sequence and classified into groups. With respect to the DNA fragment present in each group, the fragment length is measured by gel electrophoresis and used as a fingerprint. Accordingly, the present invention is advantageous in that the fragment length does not overlap with each other but a high resolution is obtained.

[0025] The present invention is applicable to various samples without much labor and time and can provide the fingerprinting method suitable for assay of many DNA or DNA fragments or analysis or assay for long DNA, as well as the analyzer or instrument in use therefor.

[0026] According to the present invention, steps of cloning and culture, which were necessary for display of mRNA for gene expression analysis in the prior art and required a long period of time, are unnecessary so that display of mRNA can be made in a shorter time which is less than one-tenth of the prior art. In addition, the data thus obtained can be converted into digital data which are preferable database for handling.

[0027] Hereinafter the present invention will be described by referring to Fig. 1. Sample 1 is digested with restriction enzyme HhaI and poly A is added to the 3' terminal base sequence four (4) of each DNA fragment. A set of sixteen (16) fluorophore tagged DNA probes 6 having a structure of RXY (wherein X and Y represent A, C, G or T) are prepared. A solution containing sample DNA composed of all of the fragments is divided into sixteen (16) fractions 15 and different DNA probes 6 are added to the respective fractions to perform complementary strand synthesis. The DNA probe hybridizes with the DNA fragment but only the DNA probe (*R-AA) which fully hybridized at the 3' terminus as in 11 is extended by complementary strand synthesis. A complementary strand 13 having the same length of the extended DNA sequence, the length of the extended complementary strand is appreciated using a fluorescent gel electrophoresis device . In gel electropherogram, the fragment length does not overlap with each other but a high resolution in length is obtained. The method and analyzer or instrument are thus applicable to various samples.

BRIEF DESCRIPTION OF THE DRAWINGS

[0028] Fig. 1 is a drawing for explaining the operational procedures in Example 1 of the present application.

[0029] Fig. 2A shows an electropherogram measured by mixed DNA fragments obtained by applying the operational procedures in Example 1 of the present invention to sample -DNA.

[0030] Fig. 2B shows an electropherogram measured by classified DNA fragments in terms of the terminal base

sequences, which are obtained by applying the operational procedures in Example 1 of the present invention to sample -DNA, using a set of sixteen (16) DNA probes.

[0031] Fig. 3A shows an electropherogram from pUC19, which is obtained in Example 2 of the present invention.

[0032] Fig. 3B shows an electropherogram from pUC118, which is obtained in Example 2 of the present invention.

[0033] Fig. 4 is a drawing for explaining the operational procedures in Example 2 of the present application.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

[0034] The present invention is applicable to polynucleotide samples such as DNA or RNA. Analysis or assay for polynucleotide samples are performed by a fingerprinting analysis or an assay using fingerprinting method.

[0035] The embodiments of the present invention will be described below in detail, with reference to the drawings.

Example 1

[0036] In Example 1, the present invention is explained, focusing on its basic principle. For DNA diagnosis, there has been generally employed a technique for digesting DNA with a restriction enzyme, fractionating the resulting DNA fragments by gel electrophoresis and perform diagnosis based on the electropherogram obtained. However, where the kind of DNA fragments produced by a restriction enzyme increases, the DNA fragments are not discriminated from each other to cause a problem in diagnosis. These fragments are characterized by the base lengths and sequences. Though each fragment has a sequence inherent thereto, the terminal sequences each consisting of several bases are also generally different from each other. For example, the digestion site by a 4-base cutter restriction enzyme generally appears at every 256 bp in average. Thus, when a double stranded DNA of approximately 50 kb is digested with a 4-base cutter restriction enzyme, about 200 double stranded DNA fragments (about 400 single stranded DNA fragments) are formed. These DNA fragments are first fractionated by the 3' terminal sequence of two bases adjacent to the recognition sequence, the number of combinations of the two base sequence becomes 4 x 4 = sixteen (16) combinations. An average number of single stranded DNA fragments belonging to each group is approximately 25 fragments. It is easy to fractionate approximately 25 DNAs having various lengths from several bases to 1 kbs. Where intact DNA is long or where a sample containing many more DNA fragments is handled, the DNA fragments are grouped using a longer base sequence. For example, where grouping is performed by the terminal 4 base sequence, the DNA fragments are fractionated into 256 groups. When the DNA fragments in each group are fractionated by length, at least 100 kinds of DNA fragments having several bases to 1000 bases can be discriminated so that more than 20,000 DNA fragments can be fractionated as a whole. The thus obtained data can be employed as fingerprints. Next, specific procedures are explained below in detail, by referring to Fig. 1.

[0037] λDNA is used as sample 1 and as a restriction enzyme HhaI is used. HhaI is a restriction enzyme which digests double stranded DNA:

5' - N ... NGCGCN ... N - 3'
3' - N ... NCGCGN ... N - 5' as shown below:
5' - N ... NGCG CN ... N - 3'
3' - N ... NC GCGN ... N - 5'

wherein N represents any one of A, T, G and C. The 3' terminal base sequence 4 of each DNA fragment 3 is GCG. in Fig. 1, two bases 10 adjacent to the 3' terminal base sequence 4 of each DNA fragment 3 is generally shown by NN, wherein N is any one of A, C, G and T.

[0038] An oligomer is introduced into the DNA fragment 3 at the 3' terminus thereof through ligation. Alternatively, poly A (which may be one base of A, C, G and T) is added to the DNA fragment at the 3' terminus thereof using a terminal deoxynucleotidyl transferase. Where poly A is added, the 3' terminus should be either 3'-protruding end or blunt end. By referring to Fig. 1, an embodiment in which poly A 5 is added is explained. When poly A 5 is added to each DNA fragment at the 3' terminus thereof, the 3' terminal sequence is represented by:

GCGAAAA ... A - 3'

Where poly A is added, the presence of recognition sequence GCG is critical for clarifying the terminal site. This is because the length of poly A cannot be controlled. Therefore, a set of sixteen (16) DNA probes 6 having a construction shown by SEQ NO. 1:

5' - TTTTTTTTTTTTTTTTTTTCGCXY - 3'

wherein X and Y each represents A, C, G or T, are prepared (the sixteen (16) DNA probes are all possible combinations of X and Y). In Fig. 1, these DNA probes are shown by: RXY, wherein X and Y each represents A, C, G or T and R represents TTTTTTTTTTTTTTTTTTTCGC. In Fig. 1, symbol * represents a fluorophore tag. Where the sequence CGC is missing, XY portion may not function for the discrimination of DNA fragments with respect to TG, TC, TT, TA, etc. (this is because XY is linked with poly T).

[0039] The sequence from the 5' end to CGC of the DNA probes is complementary to all of the DNA fragments obtained by digestion with restriction enzyme HhaI and hybridizes commonly with all of these fragments. An embodiment wherein XY (i.e., two bases 10 adjacent to the 3' terminal base sequence 4) is AA, is explained below. In this case, the XY portion hybridizes only with a fragment having complementary strand sequence TT to form double strand 11. Where the XY portion 10 has no complementary sequence to the fragment, the XY portion does not hybridize with any fragment, which is shown by numeral 12. It is not apparently appreciated only by stability of hybridization whether or not the portion XY fully hybridized. This is because almost all sites of the DNA probes hybridize commonly with any fragment but a difference in the terminal two base sequence is not sufficient for the discrimination.

[0040] However, by a complementary strand synthesis using a DNA polymerase, it can be discriminated whether or not the terminal two bases fully hybridize. A solution containing sample DNA composed of all DNA fragments obtained by digestion with restriction enzyme HhaI is divided into sixteen (16) fractions 15. Each different DNA probe 6 is added to each fraction for complementary strand synthesis. The DNA probes hybridize with the DNA fragments but only the DNA probe (*R-AA), in which the 3' end fully hybridizes, as shown by numeral 11, is extended by complementary strand extension, thereby to obtain complementary strand 13 having the same length as that of the DNA fragment. Where the DNA probes arelabelled with a fluorophore, the length of the extended complementary strand can be appreciated using a device for fluorescence gel electrophoresis.

[0041] Figs. 2A and 2B indicate a part of the results obtained using -DNA. Where it is attempted to simultaneously measure lengths of all DNA fragments . obtained by digestion with restriction enzyme HhaI, many peaks overlap with each other as shown byelectropherogram 21 in Fig. 2A, which makes discrimination impossible. However, in electropherogram 22 of the DNA fragments grouped by the terminal base sequence, using a set of sixteen (16) DNA probes 23 (in Fig. 2B, only the XY portion of the DNA probe *RXY is shown) the respective peaks corresponding to the respective fragments are separated from each other as shown in Fig. 2B. Accordingly, a length of each DNA fragment in the grouped DNA fragments can be determined. The electropherogram of each DNA fragment shown in Fig. 2B is inherent depending upon kind of intact DNA to be analyzed; if the structure of intact DNA is different, another electropherogram appears. Thus, these electropherograms can be utilized as fingerprints, which are in turn utilized for DNA diagnosis, etc. In Figs. 2A and 2B, the abscissa indicates base length 24.

Example 2

[0042] Figs. 3A and 3B indicate the results obtained by applying the present invention to the discrimination of different DNAs. The operational procedures are similar to Example 1. Two kinds of pUC having somewhat different base sequences are employed as samples, which are pUC19 and pUC118. pUC19 and pUC118 are DNA composed of 2686 bases and DNA composed of 3162, respectively, in which approximately 2630 bases give the same common sequence. After digestion with restriction enzyme HhaI, an oligomer with a known sequence at the terminus of the digested fragment is added by ligation. Thereafter in a manner similar to Example 1, the DNA fragments are grouped by the terminal base sequence of the digested fragment using a set of sixteen primers (DNA probes) and the electropherogram of each group is obtained.

[0043] The electropherograms 31 shown in Fig. 3A are those from pUC19. The electropherograms 32 shown in Fig. 3B are those from pUC118. The abscissa in Figs. 3A and 3B represents base length 24. As the fragments which are missing in the electropherograms from pUC19 but detected only in the electropherograms from pUC118, about 75 bases (CT, GT primers) and about 450 bases (CC, GT primers) are noted and as the fragment detected only in pUC19, about 130 bases (CT, GT primers) are found.

[0044] As stated above, by applying the method of the present invention, various DNAs which are slightly different from each other can be discriminated in a simple manner. Of course, the present invention is also applicable to genomic DNA. When DNA is digested with restriction enzyme HhaI, the digestion site is found once per 250 bases in average, since HhaI is a 4-base cutter restriction enzyme. In the case of genome, about 100 double stranded DNA fragments in average are produced per 1 M base. Where a 8-base cutter restriction enzyme such as NotI, etc., is employed in assaying a genome of about 100 M bases, DNA fragments of almost the same order are produced and hence, the method described hereinabove is applicable to this case as it stands.

[0045] It is extremely difficult to fractionate such a vast number of DNA fragments (2000 to 3000 fragments when calculated as single strands) by gel electrophoresis and find a slight difference in their electropherograms. Therefore, DNA fragments are classified into 16 groups in terms of the 3' terminal sequence of each DNA fragment digested with a restriction enzyme, whereby approximately 150 DNA fragments in average are contained in one group. Since this number is smaller than the number shown in Example 1, the DNA fragments can be analyzed in a manner similar to Example 1.

[0046] Complementary strand synthesis is greatly affected by whether or not the 3' terminal two bases are fully complementary to the target DNA (DNA fragment) of the DNA probes. This phenomenon is utilized to divide the formed DNA fragments into 16 groups. The DNA fragments are represented by:

$$5' - C'_1 \ldots C'_i X_n \ldots X_1 C_i \ldots C_1 - 3'$$

wherein $X_n \ldots X_1$ is a sequence inherent to each fragment; $C_i \ldots C_1$ and $C'_1 \ldots C'_i$ represent recognition sequences or a part thereof and are known. 18-mer oligomers, $A'_1 \ldots A'_{18}$, $A_{18} \ldots A_1$, are ligated with this DNA fragment at the 5' and 3' termini thereof, respectively. As the result, each DNA fragment comes to have the following structure:

$$5' - A'_1 \ldots A'_{18} C'_1 \ldots C'_i X_n \ldots X_1 C_i \ldots C_1 A_{18} \ldots A_1 - 3'$$

$A_{18} \ldots A_1$ and $A'_1 \ldots A'_{18}$ are oligomer with known sequences. In each DNA fragment, the sequence of $C_i C_1 A_{18} \ldots A_1$ is known and several bases in the 3' terminal $X_n \ldots X_1$ adjacent to the known sequence are utilized for grouping of the DNA fragments.

[0047] DNA probes which hybridize with this DNA fragment are prepared. Such DNA probes are those complementary to $A_1 \ldots A_{18} C_1 \ldots C_i$ of the DNA fragment and fully hybridize with the specific sequence of X_1X_2. Since X_1X_2 takes 16 possible combinations, sixteen (16) DNA probes are prepared. The DNA probes are represented by:

$$5' - A'_1 \ldots A'_{18} C'_1 \ldots C'_i Y_1 Y_2 - 3'$$

A DNA analogue such as inosine may be inserted between C'_i and Y_1 or in place of C'_i and the resulting probe may also be employed. This DNA probe is tagged with biotin (B); if necessary and desired, the tagged probe is designed to be isolated by a microtiter plate having streptavidin immobilized thereto or by magnetic beads, etc. In the above explanation, $C_1, \ldots, C_i$; $C'_1, \ldots, C'_i$; $X_i, \ldots, X_n$; Y_1, Y_2; $A_1, \ldots, A_{18}$; $A'_1, \ldots, A'_{18}$ represent any one of A, C, G and T; and,

sequence $C_1 \ldots C_i$ and sequence $C'_1 \ldots C'_i$; sequence $A_1 \ldots A_{18}$ and sequence $A'_1 \ldots A'_{18}$; and sequence X_1X_2 and sequence Y_1Y_2 form complementary strands as will be later shown in Fig. 4.

[0048] The present invention will be further described in detail, by referring to Fig. 4. DNA sample 38 (18-mer oligomers with known sequences: $A'_1 \ldots A'_{18}$ and $A_{18} \ldots A_1$ are connected to each DNA fragment at the 5' and 3' termini thereof, respectively; which was obtained by digesting DNA with HindIII and the thus obtained DNA fragments were grouped) is divided into 16 fractions. Each fraction is separately charged in a vessel. One biotinylated DNA probe 39 (wherein B represents biotin):

$$5' - B - A'_1 \ldots A'_{18} C'_1 \ldots C'_i Y_1 Y_2 - 3'$$

is added to each fraction. DNA polymerase and substrates for strand synthesis (dNTP; deoxynucleotide triphosphate) are further added thereto to perform a complementary strand extension reaction.

[0049] As a DNA polymerase, thermostable Taq or thermostable sequenase is employed. The reaction is carried out under conditions of thermal cycling to synthesize a complementary strand of the DNA fragment completely complementary to the terminal sequence of DNA probe. After the complementary strand synthesis, magnetic beads 42 with streptavidin (Av) on its surface are added to capture the biotin-labelled synthesized DNA strand 40 on the beads. Plastic beads or filters having streptavidin immobilized thereto may also be employed in place of magnetic beads. The non-reacted substances captured other than the DNA strand are removed. As described above, sixteen (16) groups 43 are prepared for each DNA fragment having a specific terminal sequence:

$$X_2 X_1 C_i \ldots C_1 A_{18} \ldots A_1$$

In this Example, complementary strand to each DNA fragment is employed for the grouping. By elevating a temperature of the solution containing DNA fragments, intact DNA fragments hybridized are liberated and removed. As shown in Fig. 4, the group of complementary strands (DNA strand) 46 of recovered fragment have a common probing sequence at the 5' terminus (the DNA probe sequence is located at the 5' terminus of the synthesized strand after complementary strand synthesis) and are fractionated into 16 groups, by a difference in the two bases adjacent to the probing sequence. Each group contains 150 DNA strands in average.

[0050] The DNA fragments in each group are further discriminated and classified in terms of a difference in the 3' terminal sequence (the 3' terminal sequence corresponds to the 5' terminal sequence of intact DNA fragment used as a template in the complementary strand synthesis). The DNA strand 46 captured on the beads contains an oligomer sequence $A_1 \ldots A_{18}$ connected upon ligation, with which the DNA probe can hybridize. A set of sixteen (16) fluorophore tagged probes 44 having the same sequence as described above:

$$5' - {}^{*} A'_1 \ldots A'_{18} C'_1 \ldots C'_i Y_3 Y_4 - 3'$$

wherein * is a fluorophore tag and Y_3 and Y_4 represent any one of A, C, G and T. The DNA fragments captured on the beads are divided into 16 fractions.

[0051] As in the embodiment described above, one DNA probe is added to each fraction to perform a complementary strand synthesis (extension reaction). The reaction is carried out with 16 x 16 = 256 groups and the electropherogram of the extended product 45 is assayed. The aforesaid fluorophore tag may be altered for every probe, probes labelled with fluorophores having different colors are reacted at once, and the extended product 45 may be assayed with a multicolor detection type DNA analyzer. When using fluorophores with four different colors, it is sufficient to perform the reaction 4 times for each DNA group, i.e., 4 x 16 = 64 times in total. In addition, 64 passways are sufficient for electrophoresis. In this embodiment, the oligomer connected to the terminus was regarded as the same but different restriction enzymes may also be used at the both termini so that different sequences may be connected at the termini.

[0052] Comparison in electropherograms between E. coli JM109 and C600, which genomes are slightly different, reveals that most are the same but there is a difference in a part, like for example of pUC19 and pUC118 as shown in Figs. 3A and 3B. The difference will not be notable when electrophoresed without grouping.

[0053] As stated above, after a large DNA is digested with a restriction enzyme, the DNA fragments are grouped by a sequence around the two termini and subjected to electrophoresis to obtain a pattern for length separation, whereby an electropherogram inherent to each genome can be obtained. The pattern is effective for examination of a difference in genome. In particular, the grouping by the terminal base species is effective also for constructing a database. In pattern data on two dimensional electrophoresis alone, DNA length given by analog information is uncertain so that it is difficult to construct a database. However, the grouping by the terminal base species in combination with digital information can accurately discriminate DNAs, even if the length separation by analog information is not highly accurate.

[0054] In the foregoing explanation, the temperature of the solution containing DNA fragments is elevated to isolate and remove the intact DNA fragment hybridized. The intact DNA fragment hybridized with the DNA probe extended by complementary strand synthesis may also be used after fractionation. In this case, the solution is first heated to 80-85°C to liberate and remove the DNA fragment hybridized with the probe not extended by complementary strand synthesis. The solution is then heated to 95-100°C thereby to liberate and recover the DNA fragment hybridized with the probe extended by complementary strand synthesis.

Example 3

[0055] This embodiment is to obtain information relating to display of mRNA for gene expression analysis. The number of human genes is said to be 50,000 to 100,000 species. Accordingly, 50,000 to 100,000 species of mRNAs (messenger RNA) are present. Among these mRNAs, it is assumed that approximately 3,000 mRNAs would be regularly working. It is thus important to appreciate how these mRNAs function in various tissues. In order to examine the function of mRNAs, the existing technique comprises collecting mRNA, cloning mRNA, analyzing a sequence of mRNA present in each clone and detecting frequency of displaying mRNA. However, the procedures require much labor. It is thus desired to develop a simple operation for the analysis. Analysis of cDNA complementary to mRNA has been progressed, whereby it is also possible to prepare a DNA probe capable of specifically hybridizing with each cDNA and use the DNA probe for assay. However, probes practically usable are substantially limited to approximately several tens to 100 probes. Where cDNA that cannot be detected with a probe is detected, the use of limited probes is not practical. It is thus desired to develop a system which can detect any cDNA. In view of the existing situation, the present invention is applied to display of mRNA in this embodiment.

[0056] A mRNA library (cDNA library) is prepared from tissues (J. DNA Sequencing and Mapping, 2, 137-144 (1991). In accordance with the technique described in Nature, 357, 519-520 (1992), cDNA is prepared using biotinylated poly T, mRNA-cDNA hybrid is prepared and double stranded DNA is prepared. The double stranded DNA is digested with restriction enzyme Sau3AI and an oligomer with a known sequence is connected to the digested site. HhaI (GCG↓C) or NlaIII (CATG↓) may also be employed as the restriction enzyme.

[0057] By these procedures, a DNA fragment having biotinylated poly T sequence at the 5' terminus thereof and having a recognition sequence and known oligomer at the 3' terminus is formed. Magnetic beads having streptavidin immobilized thereon are added to the DNA fragment to form biotin-streptavidin complex. The complex captures the aforesaid DNA fragment and its complementary strand. A buffer solution is added to the beads having the DNA fragment immobilized thereto, which are charged in a vessel. The resulting mixture is heated to liberate the complementary DNA fragment. The complementary DNA fragment is used as a sample. Subsequently, the sample is treated as in Example 2 to fractionate DNA fragments in terms of the 3' terminal sequence adjacent to poly A of the DNA fragments. In this case, there are 12 selective sequences adjacent to poly T, unlike Example 2. After fractionation of biotinylated DNA probe, complementary strand synthesis is performed using sixteen (16) fluorophore tagged probes. A length of the produced DNA is measured by gel electrophoresis to obtain a fingerprint.

Example 4

[0058] In order to carry out the method explained in Examples 1 to 3, a DNA analyzer comprises reaction vessels

for reacting 16 or 256 DNA probes simultaneously to obtain the complementary strand extended products, an autopipet for automatically feeding each probe to each vessel, and a device for electrophoresis. It is preferred to carry out the complementary strand extension reaction at a temperature above 65°C, since it is apparently reflected by matching or mismatching the sequence of the terminal bases whether the reaction proceeds or not. It is also necessary to compare the relative length of the products by the complementary strand extension reaction by simultaneously subjecting the extended products to electrophoresis for every base of the terminal bases of DNA probe or by subjecting a length (reference) marker to electrophoresis at the same time, for each base of the terminal bases of DNA probe. For this reason, particularly preferred is such an instrument for determining a fragment length that a reference marker is electrophoresed together with each product by the complementary strand extension reaction and relative positional relationship between the fragment of the complementary strand extension reaction product fractionated by electrophoresis and the marker is utilized. In this case, operation is easily done by discriminating the length of DNA fragments by a plurality of markers having different lengths to arrange data in such a manner as to what group the DNA fragments belong. If DNA fragments up to 1 Kbs are fractionated into 200 groups by markers appearing for every 5 bases, the DNA fragments will be even fractionated into 256 x 200 - 50,000 groups. This number gives group sufficient for diagnosis of mRNA, etc. or for fingerprinting analysis.

[0059] Measurement accuracy in electrophoresis is high in base length ranging from 10 to 500 bases. Where DNA fragments are fractionated for every 2 bases, fractionated data can be obtained in more detail. In such measurement, it is desired to label the marker with a fluorophore having quite a different emission wavelength from that of a fluorophore for labelling the target DNA fragment so that the marker can be readily discriminated from the target DNA fragment. A fluorophore has a plain emission zone toward a longer wavelength direction. It is thus important for the marker not to interfere with desired measurement of DNA fragments, by selecting a kind of fluorophore used for the marker to tag the target DNA fragment and to have a fluorescent signal in a longer wavelength region. DNA fragments may have different structures depending upon sequence length and may have different mobilities in gel electrophoresis even when they have the same base length. Even in such a case, however, the DNA fragments can be fractionated if it is appreciated in which fraction DNA fragments are displayed, using a marker as a yardstick. That is, even though DNA fragments have the same length, the DNA fragment may be fractionated in different fractions due to a difference in sequence but good reproducibility of fractionation eliminates any problem upon analysis.

SEQUENCING LISTING

[0060]

SEQ ID NO: 1

LENGTH: 24 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULAR TYPE: synthetic DNA, fluorophore tagged
SEQUENCE DESCRIPTION:

TTTTTTTTTTTTTTTTTTTCGCXY

X, Y: optional A, C, G or T

Claims

1. A method of analysis or assay for nucleotides which comprises:

(1) digesting DNA with a restriction enzyme to obtain DNA fragments;
(2) discriminating differences in sequences of 1 to 3 bases adjacent to restriction enzyme recognition sites of the DNA fragments with labeled DNA probes having all possible combinations of 1 to 3 bases out of A, C, G and T at the 3'-termini thereof;
(3) extending the labeled DNA probes by a complementary strand synthesis to fractionate the DNA fragments

into groups; and
(4) measuring the lengths of the DNA fragments which belong to said groups, or the lengths of the labeled DNA probes extended *by* said complementary strand synthesis;

wherein measured lengths of the DNA fragments fractionated into the groups by the differences in sequences of 1 to 3 bases adjacent to restriction enzyme recognition sites of the DNA fragments are employed as fingerprints.

Patentansprüche

1. Verfahren zur Analyse oder Prüfung von Nukleotiden, welches umfaßt:

(1) Verdauen von DNA mit einem Restriktionsenzym, wobei DNA-Fragmente erhalten werden;
(2) Diskriminieren der Sequenzen aus 1 bis 3 Basen, die den Restriktionsenzymerkennungsstellen der DNA-Fragmente benachbart sind, mit Hilfe von markierten DNA-Sonden, die an ihren 3'-Enden alle möglichen Kombinationen von 1 bis 3 der Basen A, C, G und T haben;
(3) Verlängern der markierten DNA-Sonden durch komplementäre Strangsynthese zur Fraktionierung der DNA Fragmente in Gruppen; und
(4) Messen der Längen der DNA-Fragmente, welche zu den Gruppen gehören, oder der Längen der durch komplementäre Strangsynthese verlängerten markierten DNA-Sonden;

wobei die gemessenen Längen der durch die Unterschiede in den Sequenzen aus 1 bis 3 Basen, die den Restriktionsenzymerkennungsstellen der DNA-Fragmente benachbart sind, in Gruppen fraktionierten DNA-Fragmente als Fingerabdrücke eingesetzt werden.

Revendications

1. Procédé d'analyse ou de dosage de nucléotides comprenant :

(1) la digestion de l'ADN à l'aide d'une enzyme de restriction pour obtenir des fragments d'ADN,
(2) la détection des différences de séquences de 1 à 3 bases adjacentes aux sites de reconnaissance d'enzyme de restriction des fragments d'ADN à l'aide de sondes d'ADN marquées ayant toutes les combinaisons possibles de 1 à 3 bases parmi A, C, G et T au niveau des extrémités 3' de celles-ci,
(3) l'élongation des sondes d'ADN marquées par une synthèse de brin complémentaire pour fractionner les fragments en groupes, et
(4) la mesure des longueurs des fragments d'ADN qui appartiennent auxdits groupes, ou les longueurs des sondes d'ADN marquées élonguées par ladite synthèse de brin complémentaire,

caractérisé en ce que des longueurs mesurées des fragments d'ADN fractionnés en groupes par les différences de séquences de 1 à 3 bases adjacentes aux sites de reconnaissance d'enzyme de restriction des fragments d'ADN sont utilisées en tant qu'empreintes digitales.

FIG.1

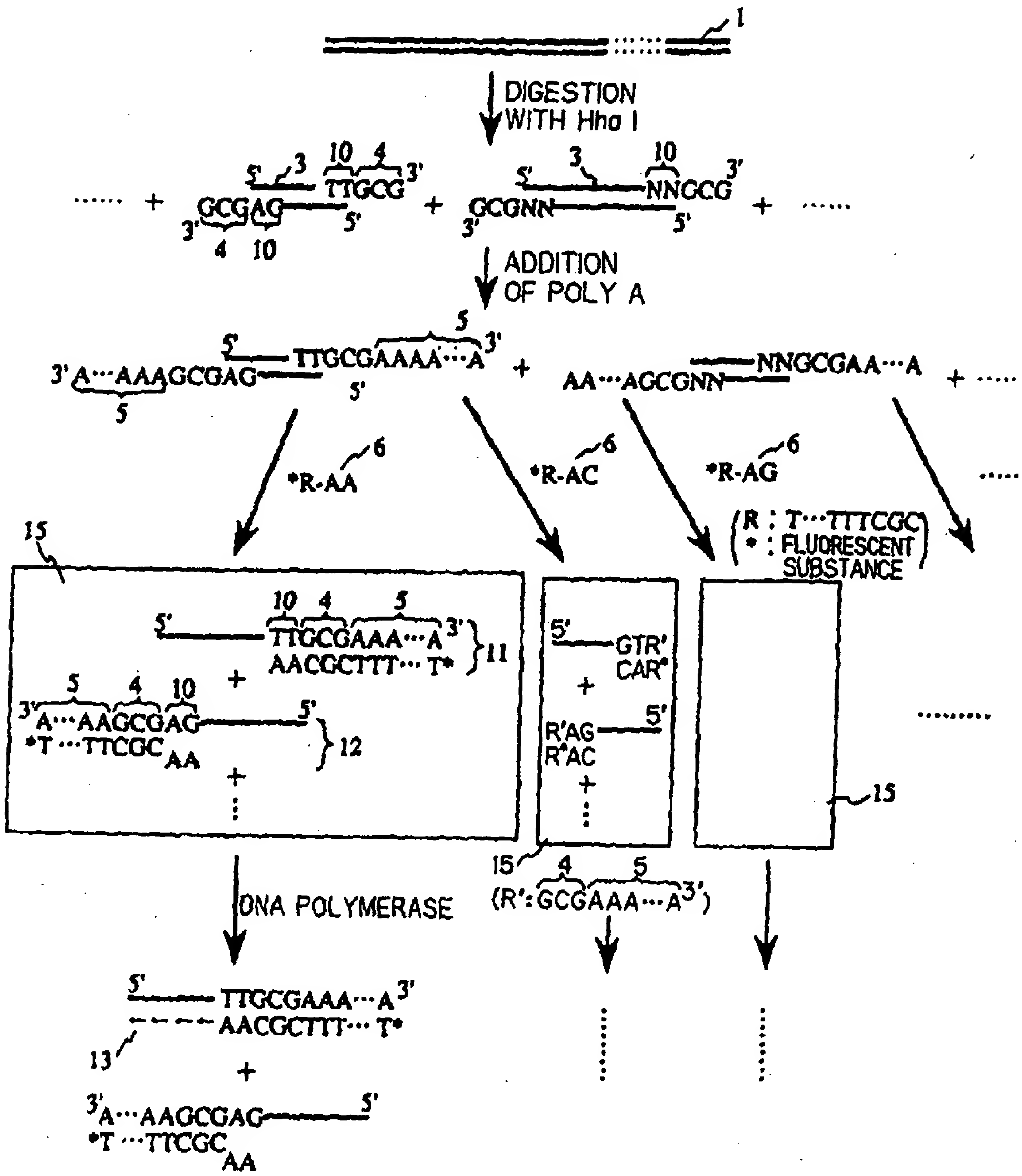

DIGESTION WITH Hha I
ADDITION OF POLY A
*R-AA
*R-AC
*R-AG
R : T···TTTCGC
* : FLUORESCENT SUBSTANCE
TTGCGAAA···A
AACGCTTT···T*
A···AAGCGAG
*T···TTCGC AA
GTR'
CAR*
R'AG
R*AC
(R': GCGAAA···A 3')
DNA POLYMERASE

FIG.2A

FIG.2B

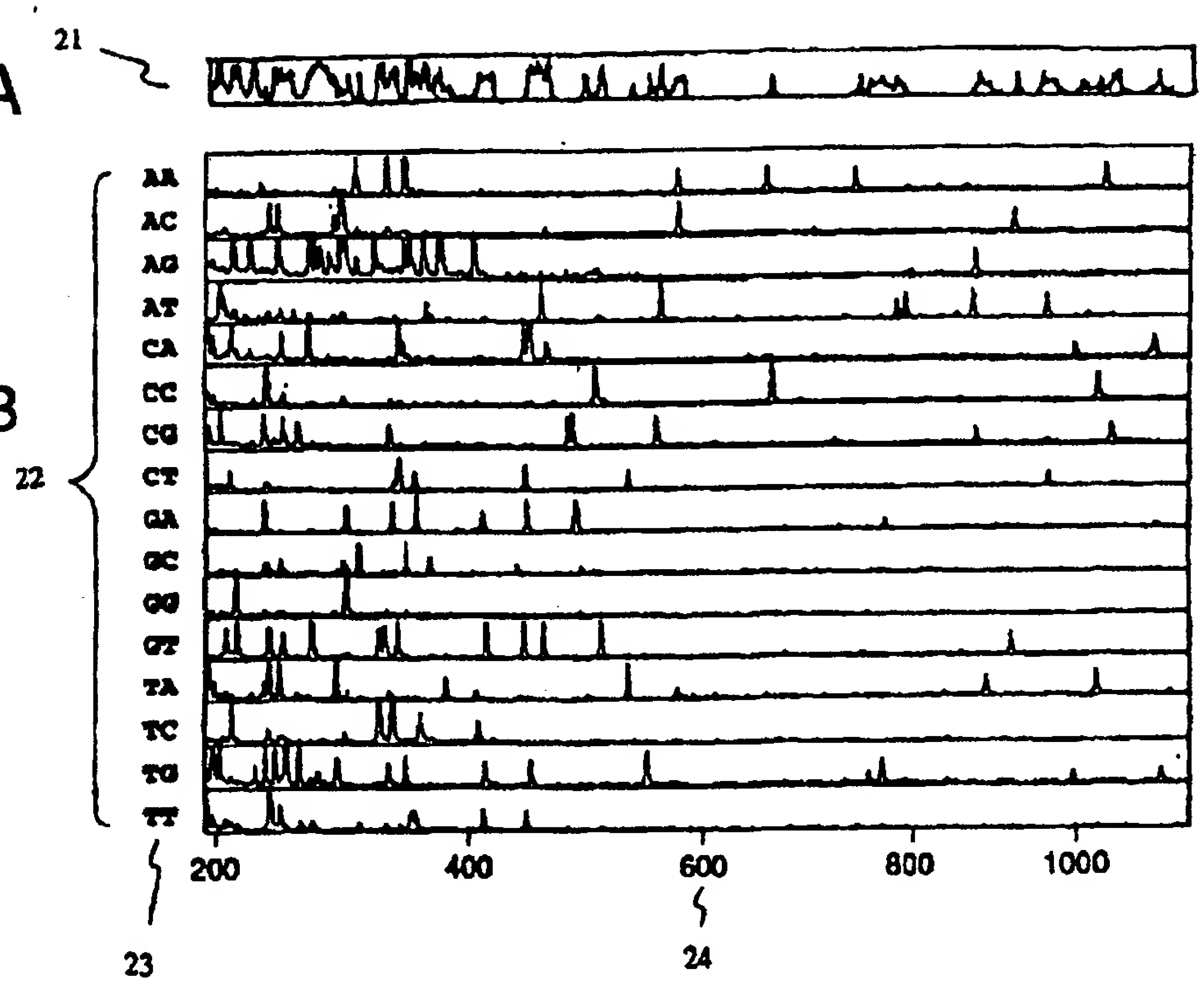

21
22
AA
AC
AG
AT
CA
CC
CG
CT
GA
GC
GG
GT
TA
TC
TG
TT
200
400
600
800
1000
23
24

FIG.3A

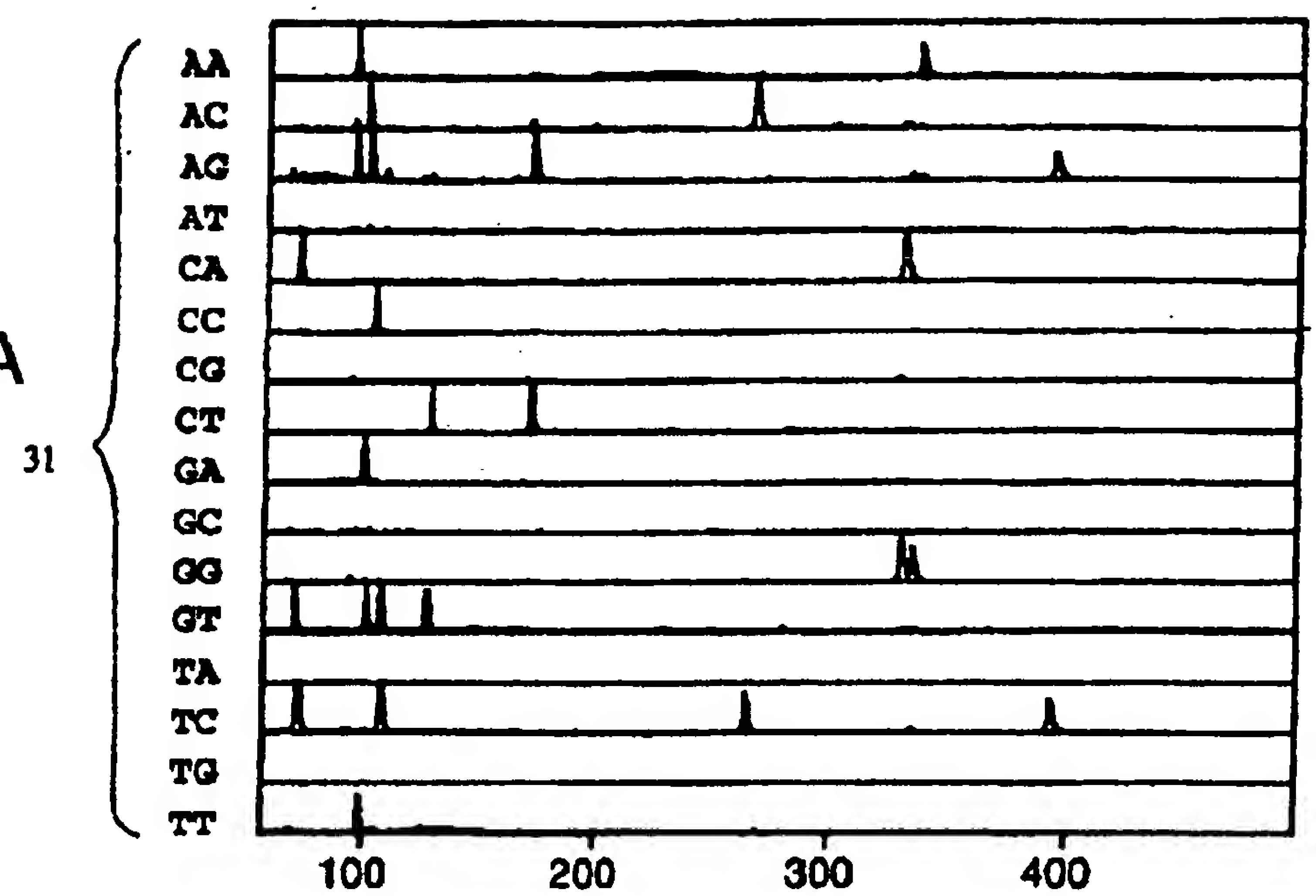

31
AA
AC
AG
AT
CA
CC
CG
CT
GA
GC
GG
GT
TA
TC
TG
TT
100
200
300
400

FIG.3B

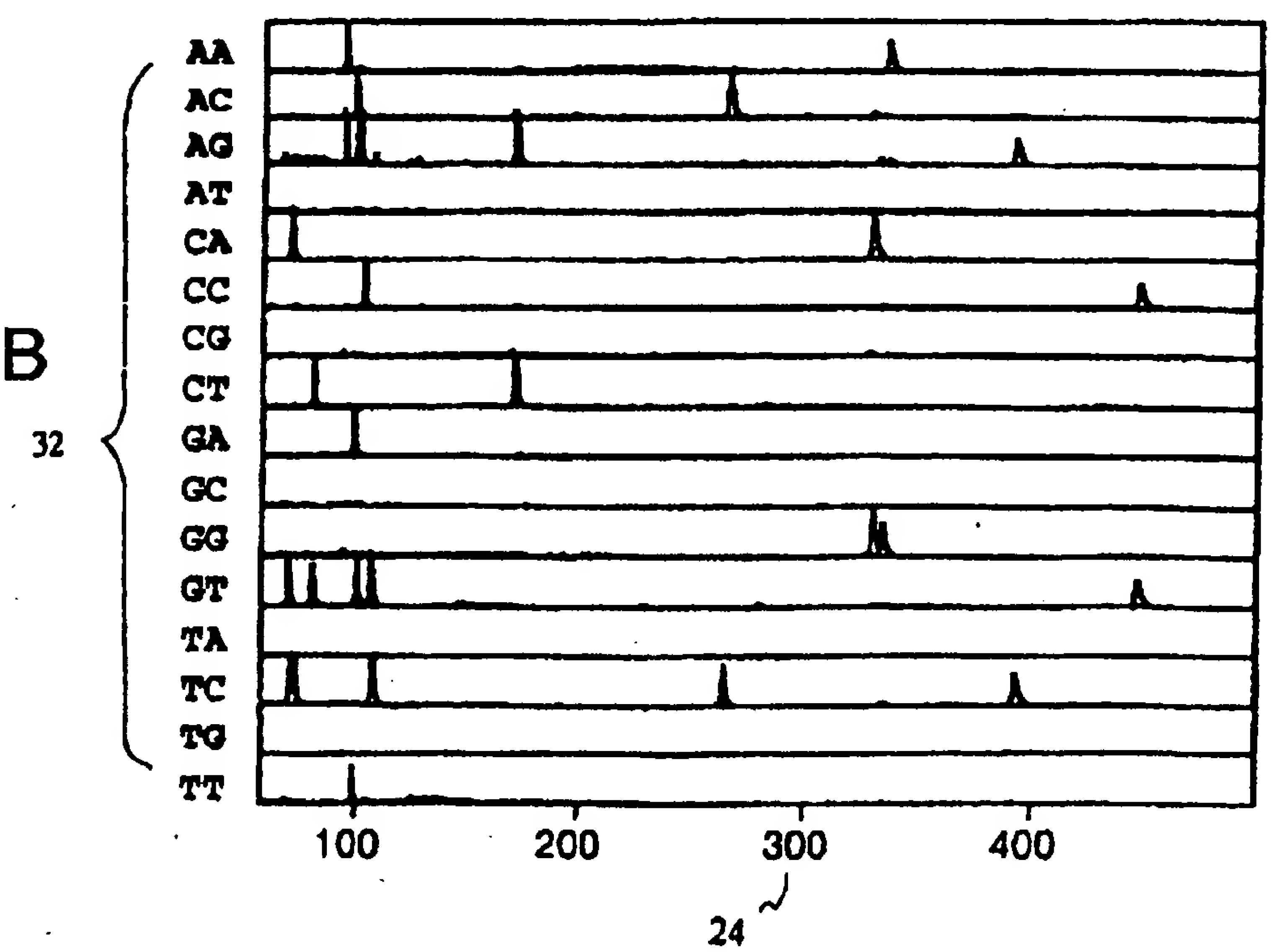
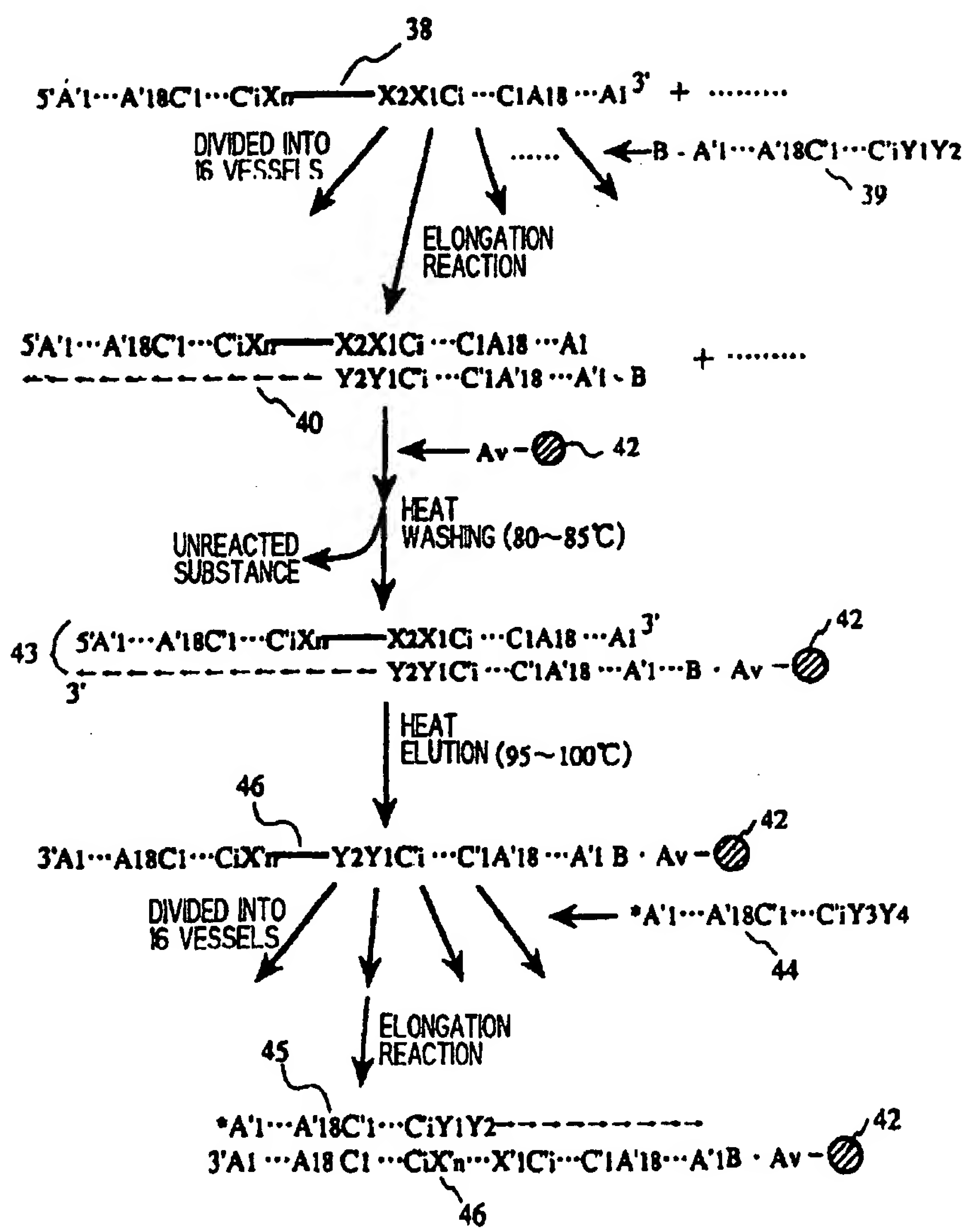

32
AA
AC
AG
AT
CA
CC
CG
CT
GA
GC
GG
GT
TA
TC
TG
TT
100
200
300
400
24

FIG.4


38
5'A'1···A'18C'1···C'iXn—X2X1Ci···C1A18···A1 3' + ·········
DIVIDED INTO 16 VESSELS
ELONGATION REACTION
B - A'1···A'18C'1···C'iY1Y2
39
5'A'1···A'18C'1···C'iXn—X2X1Ci···C1A18···A1
Y2Y1C'i···C'1A'18···A'1 - B
40
Av
42
HEAT WASHING (80~85℃)
UNREACTED SUBSTANCE
43
5'A'1···A'18C'1···C'iXn—X2X1Ci···C1A18···A1 3'
3' Y2Y1C'i···C'1A'18···A'1···B · Av
HEAT ELUTION (95~100℃)
46
3'A1···A18C1···CiX'n—Y2Y1C'i···C'1A'18···A'1 B · Av
DIVIDED INTO 16 VESSELS
*A'1···A'18C'1···C'iY3Y4
44
ELONGATION REACTION
45
*A'1···A'18C'1···C'iY1Y2
3'A1···A18 C1···CiX'n···X'1C'i···C'1A'18···A'1B · Av
46